Chordal Decomposition in Rank Minimized SDPs

Jared Miller[1], Yang Zheng[2], Biel Roig-Solvas[1], Mario Sznaier[1], Antonis Papachristodoulou[3]

[1]Northeastern University [2]Harvard University [3]University of Oxford

Motivation

Many problems in machine learning can be equivalently expressed as rank-constrained semidefinite programs (SDPs):

$$\begin{aligned} X^* = \operatorname*{argmin}_X \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i \quad i = 1..m \\ & X \succeq 0, \quad \operatorname{rank}(X) \leq t, \end{aligned} \tag{1}$$

Problem (1) is NP-hard, but the Reweighted Trace Heuristic is a convex relaxation. Rank-approximated SDPs exhibit poor scaling properties as the size of X grows.

Exploiting chordal sparsity in ML-motivated SDPs often leads to algorithms that scale linearly with the number of data points.

Chordal Graphs and Semidefinite Optimization

In sparse SDPs, only a few entries of X appear in the cost function and equality constraints. All other entries are ``free'' to choose in order to force $X \succeq 0$. If the structure is chordal, we can take advantage of it to reduce computational complexity.

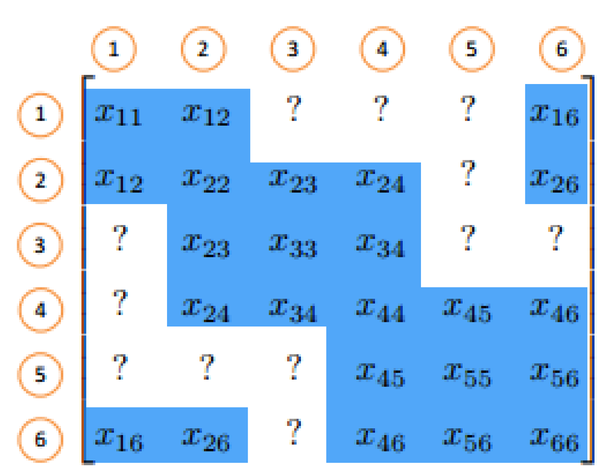

Figure 1: Variables that appear in (C, A_i) vs. `free' entries. Forms a sparsity pattern $\mathcal{G}(\mathcal{V}, \mathcal{E})$

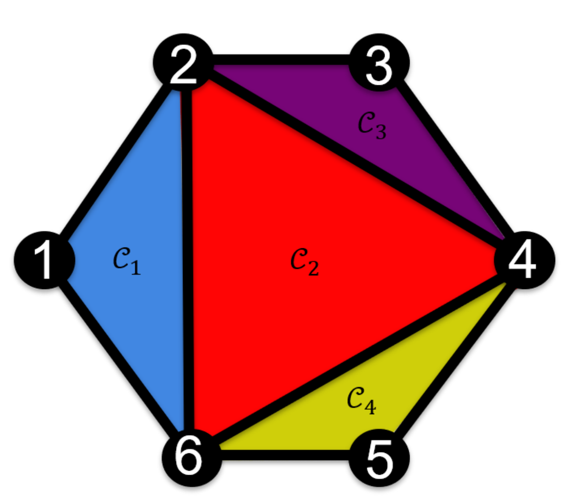

Figure 2: The chordal graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ identified with the above sparsity pattern as well as maximal cliques $\mathcal{C}_k$ (complete subgraphs). Chordal graphs have shortcuts on all 4+length cycles (triangulated).

$\mathbb{S}^n_+(\mathcal{E}, ?)$ is the set of PSD-completable matrices with sparsity $\mathcal{G}$. If all the clique submatrices $E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^T = X_k \succeq 0$ (Grone), a minimal rank completion exists:

Minimum Rank Completion

For any $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$, there exists at least one minimum rank PSD completion where

$$\operatorname{rank}(X) = \max_k \operatorname{rank}(E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^T)$$

Minimizing $\operatorname{rank}(X)$ is equivalent to minimizing the maximum $\operatorname{rank}(X_k)$.

Chordal Decomposition of Rank-Minimized SDP

The chordalized rank-minimized SDP is:

$$\begin{aligned} \min_{X, X_k} \quad & \langle C, X\rangle + \sum_{k=1}^{p} \langle W_k, X_k\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \ i = 1..m \\ & X_k = E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^T, \\ & X_k \succeq 0, \forall \ k = 1..p, \end{aligned} \tag{2}$$

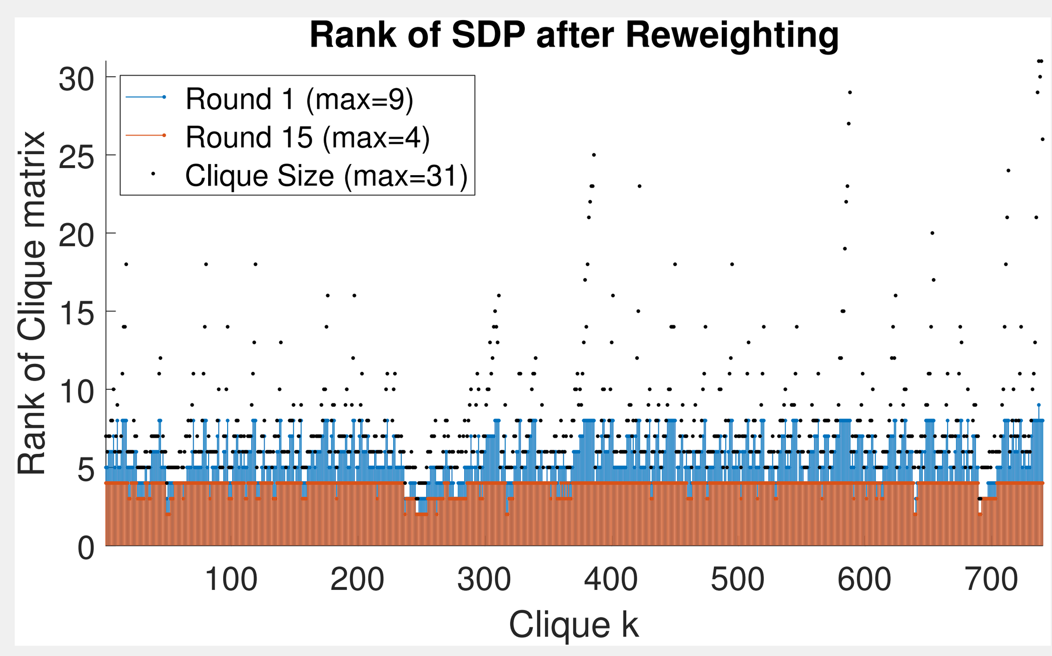

Figure 3: Chordal Rank SDP on an 1000-vertex Maxcut problem. There are 740 cliques with $|\mathcal{C}^{\max}| = 31$ (black). Maximum clique rank starts at 9 (blue), and drops to 4 (orange) after 15 rounds.

Experiments

Problem (2) is convex for each reweighting iteration. Tests were run on Subspace Clustering and Maxcut SDPs by algorithms such as interior point methods and ADMM.

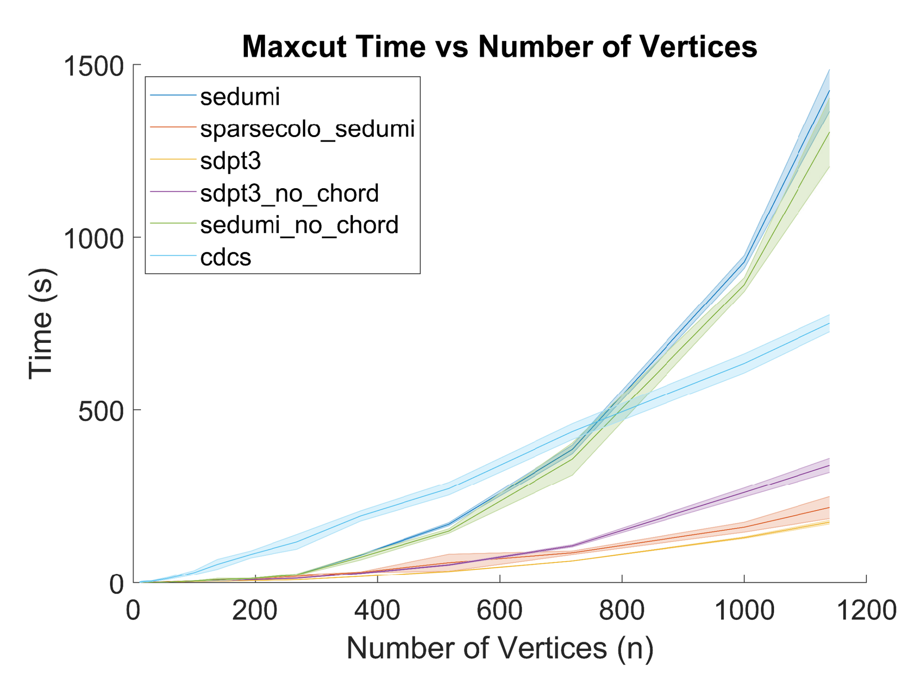

Figure 4: Maxcut time $\pm$ 1 stdev vs. $|\mathcal{V}|$, 5 trials.

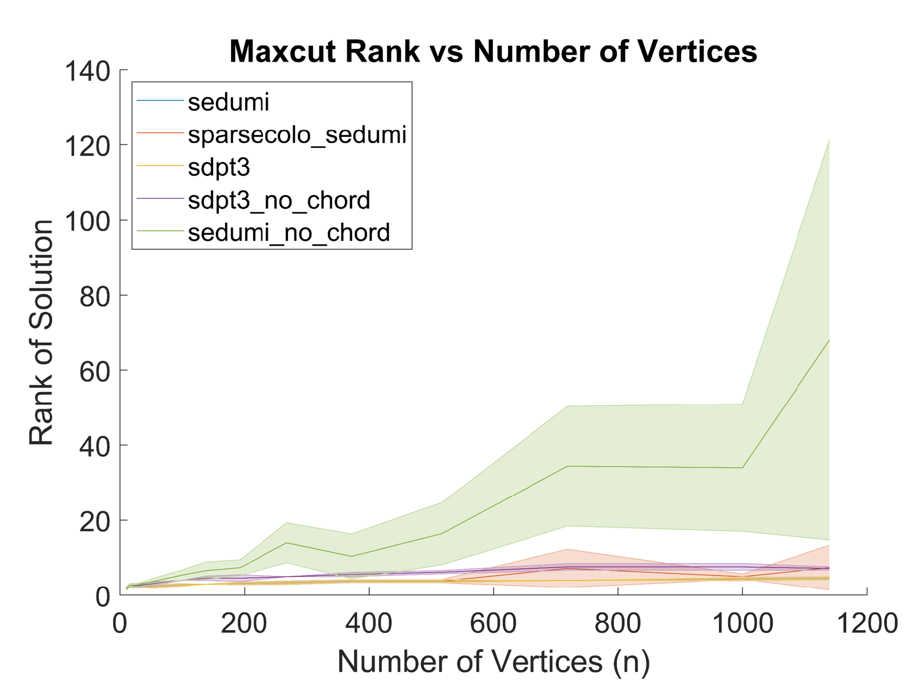

Figure 5: Maxcut rank $\pm$ 1 stdev vs. $|\mathcal{V}|$, 5 trials.

X_k can be optionally eliminated:

$$\begin{aligned} \min_X \quad & \langle C + W_{\mathcal{C}}, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, i = 1..m, \\ & X \in \mathbb{S}^n_+(\mathcal{E}, ?). \end{aligned}$$

where $W_{\mathcal{C}} = \sum_{k=1}^p E_{\mathcal{C}_k}^T W_k E_{\mathcal{C}_k}$ is the accumulated clique weight. The new cost $C + W_{\mathcal{C}}$ retains the sparsity pattern $\mathcal{E}$.

Subspace Clustering

Given N_p points $x_j \in \mathbb{R}^D$ and a N_s subspaces with unit normals $r_i \in \mathbb{R}^D$, subspace clustering aims to determine if point x_j came from subspace r_i (binary labels s_{ij}). This occurs if $r_i^T x_j = 0$, relaxed to $|r_i^T x_j| \leq \epsilon$ under bounded noise.

These algorithms allow subspace clustering to scale linearly with number of points and subspaces

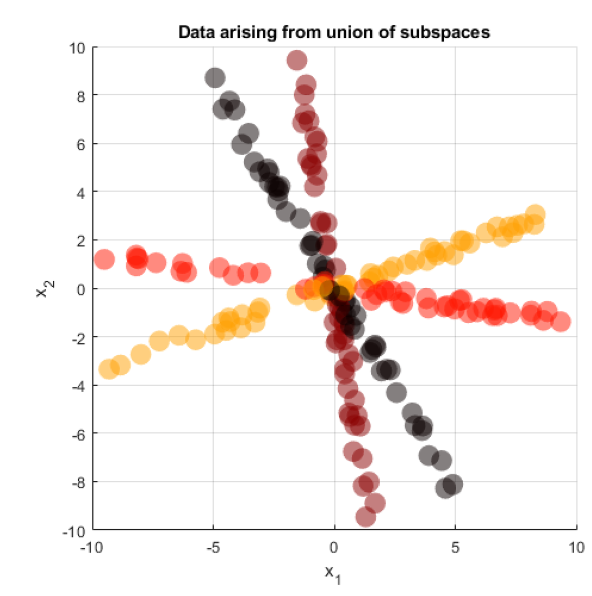

Figure 6: A typical problem in Subspace Clustering

Finding (r_i, s_{ij}) is a nonconvex problem:

$$\begin{aligned} \operatorname*{find}_{r,s} \quad & s_{ij}|r_i^T x_j| \leq s_{ij}\epsilon & s_{ij} = s_{ij}^2 \\ & \sum_{i=1}^{N_s} s_{ij} = 1 & r_i^T r_i = 1 \end{aligned} \tag{3}$$

Given $X = [1, r_i, s_{ij}][1, r_i, s_{ij}]^T$, this is a rank-1 SDP in X. **We improve Cheng *et. al.* [2016]'s chordal sparsity (grey) by using a reduced chordal extension (red).**

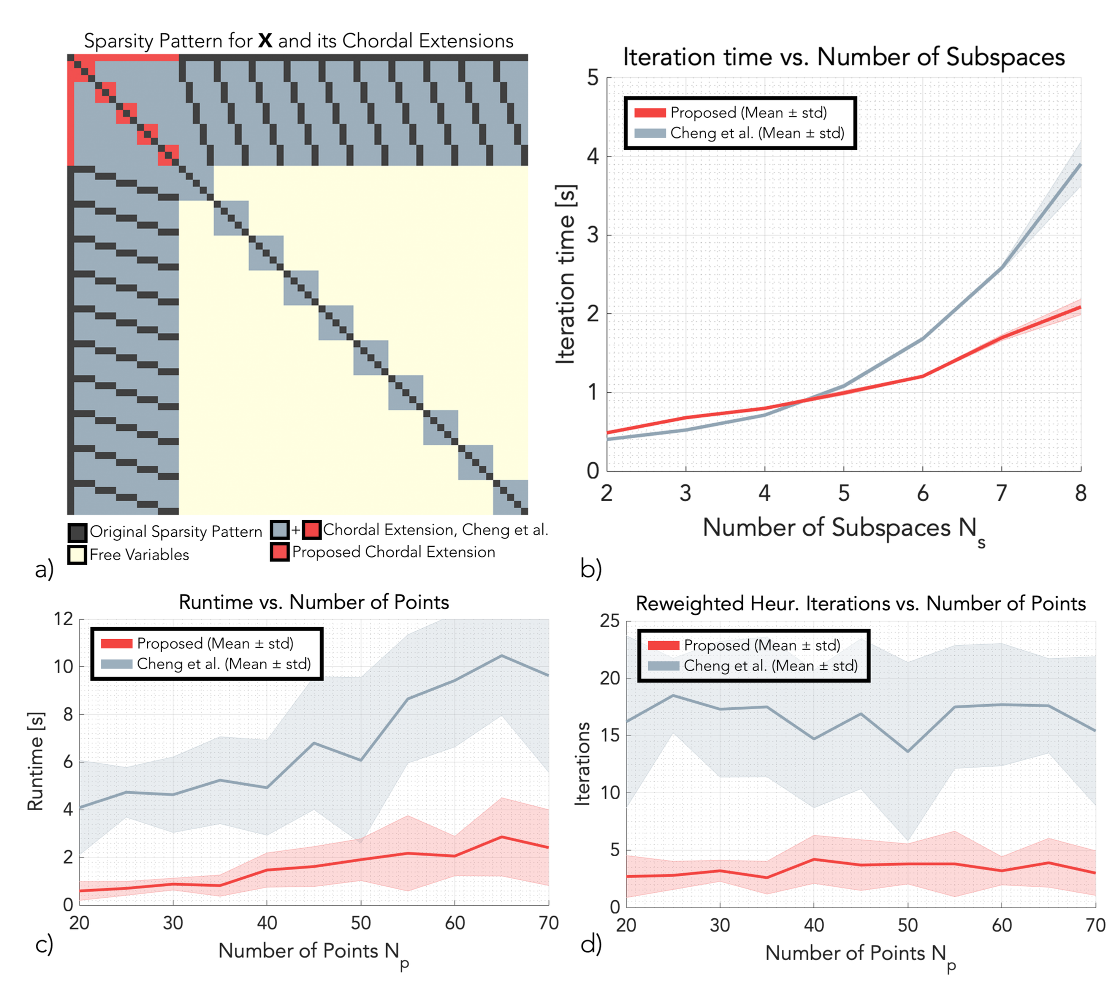

Figure 7: a) X variable structure and chordal extensions. b, c, d) Further runtime analysis.

A summary of the matrix sizes are:

Problem	Rank 1 PSD	Other PSD	%edges
Full X	$[1 + N_s(D + N_p)]$	$\varnothing$	1637%
Cheng	$[1 + N_s D]$	$N_p[1 + N_s(D + 1)]$	350%
Ours	$N_s[D + 1]$	$N_p N_s[D + 2]$	13%

Table 1: $N[k]$ denote N PSD cones: $(\mathbb{S}^k_+)^N$. %edges measures size of chordal extension over baseline $\mathcal{E}$ $(D = 3, N_p = 10, N_s = 5)$.

https://arxiv.org/abs/1904.10041